DOCUMENT RESUME

ED 119 473 FL 007 403

AUTHOR Lowe, Pardee, Jr.
TITLE The Oral Interview - A Criterion-Referenced Test?
PUB DATE Nov 75
NOTE 20p.; Paper presented at the meeting of the American Council on the Teaching of Foreign Languages (Washington, D.C., November 27-30, 1975)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage
DESCRIPTORS *Criterion Referenced Tests; Interviews; Language Instruction; Language Proficiency; *Language Tests; Linguistic Performance; Listening Comprehension; Oral Expression; Second Language Learning; *Second Languages; Speech Evaluation; *Speech Skills; *Speech Tests
IDENTIFIERS *Oral Interviews

ABSTRACT

The oral interview may be viewed as a criterion-referenced test for making either/or decisions about functional use of spoken language. Speech production can be tested by either the oral interview or the Valdis (1972) "Performance Objectives for Speaking," and dialogue between the two systems can be profitable. Current literature on criterion-referenced testing and performance objectives suggests that the major problem in previous speaking tests lay in not specifying the test's parameters. A book by Vallette and Disick and the Defense Language Institute's Handbook both suggest specifying what the task is designed to show, the nature of the task, how the task shall be tested, conditions under which the test will be taken, and criteria used to determine performance. The U.S. Government regularly conducts language proficiency tests by means of oral interviews. The Civil Service Proficiency Definitions rank ability in five levels from elementary to native or bilingual Proficiency. At the CIA Language Learning Center, additional guidelines for assigning proficiency levels and language grammar grids are also used. Guidelires cover speaking ability in subject matter and quality, as well as understanding. Oral interview tests are conducted to determine if a candidate communicates well enough in the target language to perform his job abroad, and how his performance compares with that of an educated native speaker. (CHK)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED119473

"THE ORAL INTERVIEW - A CRITERION-REFERENCED TEST?"[1]

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

PARDEE LOWE, JR.
LANGUAGE LEARNING CENTER, CIA

How do we test speaking? The honest answer is that most of us don't. To understand why, return with me to the thrilling days of yesteryear, the heyday of the language laboratory. Johannes Schmidt, Ph.D., ACTFL, AATG, sits before the master console. In front of each student are the instructions:

> "Record a five minute segment of speech on any topic covered in class or of your own choosing. Your performance will be graded on originality, accuracy of content and ability to express your thoughts in the target language."

FL007403

Half an hour later, surrounded by several piles of tapes, Johannes begins to assess the performances. Soon he is asking himself, "How did I ever get into this mess?" "What standards should I use?" "Didn't I say grammar would count?" ("No, you didn't, Mr. Schmidt!") Five hours later, Schmidt has heard all twenty students' tapes at least once. He has a migrane and flees home to four aspirin and bed with the electric blanket turned up to nine. Yet, in the middle of the night, he awakes with a

start! The comforting thought descends upon him - ETS has just produced a listening comprehension tape. "I'll grade my students on their understanding of that. Comprehension is part of speaking, isn't it?"

"No, Mr. Schmidt, it is not!" The opposite is truer: when someone's control of the target language is tested by speaking to him in it, the extent to which the student understands the question affects his ability to answer it in any reasonable fashion. For that reason, some government agencies do not give a separate understanding score on the oral interview; the assumption being that understanding must be at least the equivalent of speaking. Still, what happens to the candidate whose special job is to monitor radio broadcasts in a language in order to summarize the content of the news in English and thus never speaks a word of the target language? In such instances, it is possible to have 4-level understanding, but 0+ level speaking! - which only proves, Mr. Schmidt, that performance on a listening comprehension tape does not e 1 performance in speaking.

Yet, assuming that performance in listening comprehension indicates ability in speech production was precisely the next step most of the rest of the profession took, too. For several years, I taught an intensive course in spoken

German at a major university. Teaching assistants gave weekly grades on students' speaking performance in class. But it always disturbed me that there was never any component in either the midterms or final examination which asked the candidate to produce a ratable sample of speech.

I believe the climate is changing. With the advent of performance objectives, individualization of instruction, and criterion-referenced testing, I believe that we can construct meaningful speaking tests. If testing is conducted on the scale that Johannes Schmidt did, we are still going to have headaches, however. But the seeds of a better day were present even in Schmidt's ill-fated attempt.

Why was it doomed? The current literature on criterion-referenced testing and performance objectives suggests that the major problem lay in not specifying the test's parameters. Vallette and Disick (1972)[2] and the Defense Language Institute's Handbook (1975)[2] both suggest specifying what the task is designed to show, the nature of the task, how the task shall be tested, the conditions under which the test will be taken, and what percentage of mastery should be required of the student. Using the Valdis (1972) four-fold approach, Schmidt would now be in a position to reformulate his original instructions:

PURPOSE: "The following task is designed to test your ability to handle a realistic, unknown situation in the target language.

STUDENT BEHAVIOR: In a period of not more than three minutes, explain in a coherent narrative how you would get information on how to recover a lost suitcase at the Frankfurt airport.

CONDITIONS: Record your explanation in German on the cassette provided. No notes will be allowed.

CRITERION: You will be graded on how naturally you do this task; pronunciation, fluency, grammar and suitability of vocabulary to the task. A passing grade is performance in which 65 percent or better of the whole explanation is free from grammar, vocabulary and pronunciation errors."

Now, there is a ray of hope. The instructor has specified the task to a point where both he and the students have a fuller understanding of it. Also, because every student is assigned the same task, the instructor can compare performances. If his standards fail him (and they may because they still must be worked out in detail), he can at least place the performances in a series with the best performance at one end and the worst at the other. He can then select a cut-off point and separate the sheep from the goats.

Familiarity with Valdis (1972: 152-3) will allow him to apply their "External Standards for Speaking". The

standards given are particularly useful for testing micro-segments of learning at lower ends of the language learning spectrum. What happens, however, when an instructor wishes to test a student's overall ability in the target language for some macrosegment of learning; at the end of the high school course, at the end of the basic college course, at the end of an undergraduate major, or at the end of graduate school.

A few years ago, it was relatively unknown that the United States Government regularly conducted such testing by means of an oral interview. Recently, several articles and books have referred to the Government test: Clark (1972); Jones (1975) and Jones (To Appear); Weinstein (1975); Wilds (1975) and the DLI Handbook (1975). The oral interview started in the 1950's while criterion-referenced testing appears about a decade later. Although the Proficiency Definitions for the oral interview have been revised several times, I have been unable so far to establish any cross-fertilization. It appears to be an example of polygenesis.

The Government required a test to determine how Foreign Service Officers would perform their jobs abroad using the target language. Thus in the mid 1950's, the Proficiency Definitions and the oral interview were devised at the Foreign

Service Institute. This type of testing is now used not only in the Foreign Service Institute, but in the Central Intelligence Agency, the Defense Language Institute, the Peace Corps and for SHAPE. In the mid 1960's the definitions were accepted by the Civil Service Commission for the whole Government.

TABLE I

CIVIL SERVICE PROFICIENCY DEFINITIONS

Elementary Proficiency

S-1 Able to satisfy routine travel needs and minimum courtesy requirements. Can ask and answer questions on topics very familiar to him; within the scope of his very limited language experience can understand simple questions and statements, allowing for slowed speech, repetition, or paraphrase; speaking vocabulary inadequate to express anything but the most elementary needs; errors in pronunciation and grammar are frequent, but can be understood by a native speaker used to dealing with foreigners attempting to speak his language; while topics which are "very familiar" and elementary needs vary considerably from individual to individual, any person at the S-1 level should be able to order a simple meal, ask for shelter or lodging, ask and give simple directions, make purchases, and tell time.

Limited Working Proficiency

S-2 Able to satisfy routine social demands and limited work requirements. Can handle with confidence but not with facility most social situations, including introductions and casual conversations about current events as well as work, family, and autobiographical information; can handle limited work requirements, needing help in handling any complications or difficulties; can get the gist of most conversations on non-technical subjects (i.e., topics which require no specialized knowledge) and has a speaking vocabulary sufficient to express himself simply with some circumlocutions; accent, though often quite faulty, is intelligible; can usually handle elementary constructions quite accurately but does not have thorough or confident control of the grammar.

Minimal Professional Proficiency

S-3 Able to speak the language with sufficient structural accuracy and vocabulary to participate in most formal and informal conversations on practical, social, and professional topics. Can discuss particular interests and

special fields of competence with reasonable ease; comprehension is quite complete for a normal rate of speech; vocabulary is broad enough that he rarely has to grope for a word; accent may be obviously foreign; control of grammar good; errors never interfere with understanding and rarely disturb the native speaker.

Full Professional Proficiency

S-4 Able to use the language fluently and accurately on all levels normally pertinent to professional needs. Can understand and participate in any conversation within the range of his experience with a high degree of fluency and precision of vocabulary; would rarely be taken for a native speaker, but can respond appropriately even in unfamiliar situations; errors of pronunciation and grammar quite rare; can handle informal interpreting from and into the language.

Native or Bilingual Proficiency

S-5 Speaking proficiency equivalent to that of an educated native speaker. Has complete fluency in the language such that his speech on all levels is fully accepted by educated native speakers in all of its features, including breadth of vocabulary and idiom, colloquialisms, and pertinent cultural references.

Although not employing the same language as the Valdis (1972) Performance Objectives, the definitions describe tasks in a similar fashion. Because the most desirable level in government work is Level 3 (often designated "minimal professional proficiency), I direct your attention particularly to its wording in Table I.

TABLE I: PROFICIENCY DEFINITIONS

At the Language Learning Center, much of the material for performance objectives not contained in the Proficiency Definitions themselves are provided by two other documents: "Some Guidelines for Assigning Language Proficiency Levels" (see Table II) and a Grammar Grid for each language. The Proficiency Definitions and the Guidelines For Rating apply to all languages while Grammar Grids are language-specific in those languages where they are available. These documents attempt to characterize both in general terms and in specific grammatical terms for a given language the domains of behavior and tasks at each level.[3] Finally, these documents suggest how representative outcomes should be rated.

TABLE II: GUIDELINES

TABLE II

LANGUAGE LEARNING CENTER, CIA

FOR EXAMINERS ONLY

FOR EXAMINERS ONLY

GUIDELINES FOR ASSIGNING LANGUAGE PROFICIENCY LEVELS

Speaking - Understanding

LEVEL 0

SPEAKING: The examinee has no practical speaking proficiency. He may have a few isolated words and phrases which are of no practical use. UNDERSTANDING: The examinee understands some isolated words and phrases, but is unable to participate even in a very simple conversation.

LEVEL 1

SPEAKING - Subject Matter: The examinee has the minimum proficiency for survival on a day-to-day basis in the target country, i.e., function in simple question-and-answer situations. He knows enough at this level to satisfy ordinary courtesy requirements. He is able to ask and answer questions relating to situations of simple daily life and routine travel abroad. The examinee is also able to handle requests for services such as renting a hotel room and ordering a simple meal. SPEAKING - Quality: The examinee at this level normally makes errors even in structures which are quite simple and common. His vocabulary is limited to the type of situations mentioned above, and even in these situations he sometimes uses the wrong word. Although his pronunciation may be poor, he makes the minimum contrastive distinctions, including stress, intonation and tone patterns, necessary to make himself understood. UNDERSTANDING: The examinee is able to understand simple questions and statements relating to simple transactions involved in situations of daily life and independent travel abroad, allowing for slowed speech with considerable repetition or paraphrasing.

LEVEL 2

SPEAKING - Subject Matter: The examinee is able to talk in some detail about concrete subjects such as his own background, his family, travel experiences, recreational activities, educational background, and familiar places. SPEAKING - Quality: The

examinee has enough control of the language to be able to join sentences in limited discourse. He has good control of the morphology of the language (inflected languages), and of the most frequently used syntactical structures. Although his vocabulary is sufficient to talk with confidence about the type of topics described above, his limited vocabulary fairly often reduces him to verbal groping, or to momentary silence. A foreign intonation and rhythm may still be dominant. UNDERSTANDING: The examinee is able to comprehend questions and statements relating to common social topics, when the language is spoken at normal conversational speed. He can get the gist of casual conversations with educated or well-informed native speakers talking about subjects on the level of current events, allowing for occasional repetitions or paraphrased statements.

LEVEL 3

SPEAKING - Subject Matter: The examinee is able to converse and express his opinions about such topics as current events, including political and social problems of general nature. SPEAKING - Quality: The examinee has good control of grammar, though he makes occasional errors in low-frequency structures and in the most complex frequent structures. His vocabulary is broad enough so that he rarely has to grope for words in discussing the topics mentioned above. A foreign phonology, though apparent, is no longer dominant. UNDERSTANDING: The examinee can comprehend most of what is said at a normal conversational rate of speech. A person at this level is able to understand to a high degree more complex formal discourse, i.e., subjects on the level of panel discussions, news programs, etc.

LEVEL 4

SPEAKING - Subject Matter: Although the subject matter that the examinee is able to handle at this level may not differ very much from that of level 3, he is able to use the language in all non-technical situations, and can express his opinions almost as he would in his native language (assuming that he is "5" in his native language). The examinee is able to tailor his speech to his audience. He has nearly perfect grammar and speaks the language with extensive and precise vocabulary. Although the examinee may still have an accent, he very rarely mispronounces the language. UNDERSTANDING: The examinee can understand the content of all conversations and formal presentations within the range of his experience. With the exception of dialect variations and colloquialisms outside the range of his experience, he understands the type of language heard in speeches sprinkled with idioms and stylistic embellishments.

LEVEL 5

SPEAKING: The examinee is able to use the language in all conceivable non-technical subjects in a manner equivalent to that of an educated or well-informed native speaker of the language. UNDERSTANDING: The examinee is able to understand all types of formal and informal speech in a manner equivalent to that of an educated or well-informed native speaker of the language, including a wide range of dialect variations, colloquialisms, and cultural references.

The system works as follows: a government employee claims to speak French. Slated for a language-designated position, he is scheduled for a test. At the time of the interview, he is introduced to the two testers, and they proceed to have a chat. Yet, the oral test is a probing "conversational interview", in which the candidate's control of general language is put to the test by a series of standardized elicitation techniques. While not a direct test in the sense that a special section tests vocabulary, another grammar, etc. by its end (10-30 minutes), the testers can assign the candidate a rating on the scale of 0 (for no practical ability to communicate in the language) to 5 (for performance like that of an educated native speaker). The scale provides plusses for unusually strong performance at a given level so that the resultant scale furnishes 11 distinctions. All tests are taped for verification by a third rater.

The test is conducted with one basic question in mind:

> "Can the candidate communicate well enough in the target language to perform his job abroad?"

His performance is rated on:

> "How close does the performance come to that of an educated native speaker in the same topic(s) and/or situation(s)?"

In the oral interview, candidates are not compared to a set of norms nor are they assigned scores comparing them-

selves to one another in the norm-referenced sense with one receiving 95, the next 92, etc. Like criterion-referenced testing, the system is used for making either/or decisions: take job/don't take job; continue training/discontinue training; place in class which has already started/do not place in such a class. These are the kinds of decisions which the government has been making for over 20 years based on the oral interview and doing with a high degree of success.

Yet, the government test differs in two important respects from the other kinds of criterion-referenced testing described in most of the literature:

First, the oral interview is a proficiency test, not an achievement test. The question is not how much did the candidate learn of Chapter 4 of his Spanish text, rather how well does the candidate speak the target language compared to the performance of an "educated native speaker"? "Educated" does not mean aesthetic-literary education, but acquaintance with how his language works through schooling equivalent to lycée, liceo, Gymnasium etc. in Europe or a four-year liberal arts education in the USA.

Second, the oral interview normally tests candidates at a higher level than the performance objectives given in Valdis (1972) and their attendant examples.

A comparison of the two systems reveals that the oral interview falls mainly in Stage 4 of the Valdis (1972) system. Its position there is based on the major criterion of the proficiency definitions: "How well does the candidate communicate in the language?" The ability to communicate[4] plays a central role even at the lowest levels. A stage by stage comparison shows that Valdis (1972), Stages 1 & 2: Mechanical Skills & Knowledge respectively, dominate at the lowest Levels 0^+ and 1 of the government's Proficiency Definitions; Valdis (1972) Stage 3: Transfer is an integral part of Level 1; while Valdis (1972) Stage 4: Communication as defined by their examples, is an essential part of Levels 1^+-3 with Valdis (1972) Stage 5: Criticism part and parcel of Levels 4 & 5 in the Proficiency Definitions.

In all fairness to the Valdis (1972) system, my understanding of it derives less from the wording of their performance objectives, although such statements are infinitely more informative than the vague verbiage of yore, than from their examples. Frankly, there are similar problems in interpreting the government's Proficiency Definitions. Thus, government testers must be trained in how to use the system, what the standards mean, and how to elicit a ratable sample of speech from the candidate. After training, the standards

are best preserved if the testers test frequently (in our case, one or more times a week), and if when they are in doubt about a given test, they return to check the definitions and listen to sample calibrated tapes as examples of the levels they may be dealing with.

The rating contains a subjective element. This is partially controllable by the measures cited above and by the fact that raters at the Language Learning Center test in pairs and rate individually on the overall impression they have of the candidate's speech in the target language. The Proficiency Definitions furnish the basic parameters at each level. But few tests are paradigmatic examples of a given level. The challenge in testing lies in balancing the various possible combinations, for example, bad pronunciation with good grammar and limited vocabulary. Basically, the definitions are functional. A grammar grid may specify some control of subjunctive in German at Level 2^+, but a candidate may deal with 2^+ level topics with adequate grammar and never use a subjunctive. Assuming that all other structures, pronunciation, fluency and vocabulary speak for his being rated at Level 2^+, he will receive that rating. It is likely to be a less strong 2^+ than with the subjunctive, but it will be a 2^+. As Jones (To Appear) has rightly pointed out,

all of the subjective elements in testing cannot be eradicated. However, our task is to minimize them.

Space scarcely suffices to mention the basic issues, let alone tackle the substance of two different systems. Still, I hope to have shown the main ways in which the oral interview may be viewed as a criterion-referenced test with a major criterion against which all performance can be judged, non-norm-referenced rating, and either/or decisions about functional use of spoken language. Further, I hope to have shown that it is possible to test speech production by either the oral interview or the Valdis (1972) "Performance Objectives For Speaking" and, finally, that there can be a profitable dialogue between the two systems.

***

We left poor Johannes Schmidt huddling in his bed, totally dependent on a listening comprehension tape for testing speech production. It is time to tell him

"He! Johannes, schlaf nicht!"

Bibliography

Clark, John L. D. 1972. Foreign Language Testing: Theory & Practice. Philadelphia: Center for Curriculum Development. Pp. 121-129 on the "FSI Language Proficiency Interview".

Defense Language Institute. English Language Center. 1975. Handbook for Conducting Task Analyses and Developing Criterion-Referenced Tests of Language Skills. /San Antonio./ March 1975.

Jones, Randall L. 1975. "Testing Speaking Proficiency" in Testing Language Proficiency ed. Randall L. Jones & Bernard Spolsky. Arlington, VA: Center for Applied Linguistics. Pp. 1-7.

_____, (To Appear). "Achieving Objectivity in Subjective Language Tests" in The Proceedings of the Fourth International Congress of Applied Linguistics. Stuttgart.

Valette, Rebecca M. and Renée S. Disick. 1972. Modern Language Performance Objectives and Individualization: a Handbook. NY: Harcourt, Brace, Jovanovich. Particularly, pp. 152-158.

Weinstein, Allen I. 1975. "Foreign Language Majors: The Washington Perspective." ADFL Bulletin. 6 (May 1975): 18-27.

Wilds, Claudia P. 1975. "The Oral Interview Test" in Testing Language Proficiency ed. Randall L. Jones & Bernard Spolsky, Arlington, VA: Center For Applied Linguistics. pp. 29-38.

FOOTNOTES TO "THE ORAL INTERVIEW"

[1] The following paper was delivered in the Section on "Foreign Language Testing: A Time for New Directions" at the 1975 FIPLV/AATG/ACTFL Convention in Washington, D.C. in November 1975. It has been revised slightly in the interim.

[2] Hereafter referred to as Valdis (1975) and the DLI Handbook (1975) respectively.

[3] Similarly, Valdis (1972) begins with a taxonomy delimiting the field: then, specifies speaking in terms of Performance Objectives at the Five Stages; and finally, provides illustrative examples.

[4] Two levels in the Proficiency Definitions make explicit that communication in this sense is strongly tempered by grammatical accuracy: Level 5 demands the grammatical accuracy of a high-level diplomat while Level 3 requires consistent accuracy with a few errors permitted in the "core grammar" of the target language and a larger number of errors in less frequent structures if they are used at all. Testers generally allow for the primacy of grammar by weighting grammar more heavily than vocabulary and both of the above more than either pronunciation or fluency (see Wilds 1975: 32).